MEMO

ARCADIS Design & Consultancy for natural and built assets

To:
Cliff Firstenberg
Carlie Thompson

Copies:
Timothy Iannuzzi
Jamie Combes

Arcadis U.S., Inc.
Annapolis, Maryland 21401
Tel 443 699 6369

From:
Jacqueline Iannuzzi

Date:
November 14, 2016

Arcadis Project No.:
B0009989.0050

Subject:
Approach to Conducting the Sediment Quality Triad Assessment for the Newark Bay Study Area

---

On behalf of Tierra Solutions, Inc. (Tierra), Arcadis U.S., Inc. (Arcadis) will prepare a Draft Baseline Ecological Risk Assessment Report (Draft BERA Report) for the Newark Bay Study Area (NBSA) in 2017 as part of the ongoing remedial investigation (RI)/feasibility study (FS) process pursuant to the Administrative Order on Consent, United States Environmental Protection Agency (USEPA) Index No. CERCLA-02-2004-2010. The baseline ecological risk assessment (BERA), initiated in 2011, is being conducted in a stepwise process under the RI/FS, as described below. This memorandum describes the systematic approach and data analysis methods that will be used to conduct the sediment quality triad (SQT) assessment for the NBSA in support of the risk characterization for benthic invertebrates in the NBSA BERA.

1. BACKGROUND

The NBSA BERA process began with a 2-day workshop hosted by the USEPA Region 2 in Edison, New Jersey on June 28 and 29, 2011. Representatives from the USEPA and its consultants, Tierra and its consultants, and various federal and state regulatory agencies participated in the workshop. The workshop outcome is summarized in meeting minutes/notes developed by Tierra (2011) and approved by the USEPA.

Following the workshop, Tierra produced the Problem Formulation Document (PFD; Tierra 2013). The goal of the PFD (Tierra 2013) was to "establish the overall goals, breadth, and focus of the baseline ecological and human health risk assessments and to define the questions that need to be addressed during these evaluations." From an ecological risk standpoint, the objectives of the PFD (Tierra 2013) were to:

- Compile and summarize the relevantavailable information (at the time) for the NBSA
- Develop an ecological conceptual site model for the NBSA
- Conduct a conservative screeninglevel ecological risk assessmentto determine which chemical constituents would likely be evaluated in the BERA
- Select receptors for the BERA and develop risk questions, assessment endpoints (AEs), and measurement endpoints (MEs) for these receptors.

Following completion of the PFD (Tierra 2013), Tierra and the USEPA determined the data needs and scopes of work for data collection for the BERA. As part of this process, Tierra conducted an ecological field reconnaissance of the NBSA (Tierra 2015a) to refine the understanding of the habitats and potential ecological receptor use of the NBSA environs, as well as to help select appropriate sampling locations for the BERA data collections.

The data needs for the BERA and basis for sampling (sample type, numbers, and locations) are summarized in a risk assessment scoping memorandum (Tierra 2015b) that was developed in an iterative manner by Tierra and the USEPA between 2013 and 2015. The scoping memorandum (Tierra 2015b) contains an updated version of the original table of the AEs and MEs for the NBSA BERAfrom the PFD (Tierra 2013).

The BERA sampling program was implemented in stages between 2014 and 2016 and is now complete. It included three primary sampling programs:

1. Clam, crab, and co-located surface sediment sampling
2. Sediment quality triad (SQT; synoptic data on surface sediment chemistry, sediment toxicity, benthic communities, and invertebrate bioaccumulation) and porewater sampling
3. Fish tissue and community sampling.

The specific sampling programs are described in a series of Quality Assurance Project Plans (Tierra 2014a, 2014b, 2015c). The field investigation and data results from the BERA sampling are summarized in a series of draft reports:

- Clam and crab field investigationand data reports (Tierra 2016a, 2016b)
- SQT and porewater field investigationand data reports (Tierra 2016c, 2016d)
- Fish field investigationand data reports (Tierra 2016a, 2016e).

These BERA datasets, in conjunction with the sediment chemistry data from Phases I and II of the RI and data presently being collected under Phase III of the RI, will constitute the site-specific data to be used to conduct the risk assessments (i.e., both ecological and human health).

2. OBJECTIVES AND APPROACH

This memorandum describes the systematic approach and data analysis methods that will be used to evaluate the sediment and porewater chemistry data, sediment toxicity data (i.e., *L. plumulosus* bioassay

results), benthic invertebrate community (BIC) data, and bioaccumulation data (*N. virens* bioaccumulation test results) collected as described and reported by Tierra (2015c, 2016c, 2016d). Collectively, the analysis of these datasets constitutes the SQT assessment that, in conjunction with the tissue-based exposure/effects assessment to be conducted for infaunal invertebrates and epibenthic clam and crab data, will constitute the risk assessment for benthic invertebrates in the NBSA BERA. The latter assessment will be described in a technical memorandum being developed by Arcadis that describes the preliminary exposure factors and toxicity reference values (TRVs) that will be used for the tissue-based risk characterization in the BERA.

A major influence on the analysis of the SQT data for the LPRRP was the data analysis methods imposed on the Cooperating Parties Group (CPG) by the USEPA and described in the following documents:

- *Lower Passaic River Study Area Revised Baseline Ecological Risk Assessment* (Windward 2016)
- *SQT Statistical Guidance Technical Memorandum* (USEPA and Battelle 2015)
- Final Comment 71 to the Draft BERA (re. Sediment Quality Triad Methodology) prepared by USEPA (USEPA 2015).

These methods were incorporated in the current plan for analysis. The SQT methodology (USEPA 2015) was incorporated as closely as possible with a few exceptions related to differences in data collection. Specifically, the toxicity scoring was adapted as appropriate given the differences in the organism tested and the measured endpoints and the sediment chemistry scoring was revised to include the porewater component with equal weighting. The statistical guidance (USEPA and Battelle 2015) provided a basis for the methods proposed for the statistical evaluation.

The SQT and porewater data were collected from 30 stations in Newark Bay in 2015. At each station, samples were collected for analysis of surface sediment chemistry, surface sediment porewater chemistry, sediment toxicity testing (i.e., *L. plumulosus* laboratory bioassay), and BIC taxonomy. The following steps will be followed in the analysis of these data:

1. Data reduction will be conducted to develop a list of chemicals that may be impacting benthic invertebrate communities in the NBSA (Section 3).
2. An SQT evaluation—an objective method of comparing sediment chemistry concentrations, sediment toxicity endpoints, and BIC metrics to screening, control, or reference values—will be conducted to categorize each sampling station with respect to the degree of potential impact (Section 4).
3. Statistical modeling, including univariate and multivariate analyses, will be conducted to evaluate the association of sediment and porewater chemistry with effects measures (i.e., toxicity endpoints and BIC metrics) (Section 5).

Sediment from eight of the SQT stations were also selected for a laboratory bioaccumulation study with the polychaete, *N. virens*. The data from this study will be evaluated using linear regression to quantify a predictive relationship between sediment and tissue concentrations, where found.

Each of the aforementioned steps/evaluations are described in detail below.

3. DATA REDUCTION METHODS

3.1 Chemistry Data

Sediment chemistry, reported on a dry weight (dw) basis, will be used in the statistical and SQT evaluations. Normalization to total organic carbon (TOC) may be used in the statistical evaluation for organic chemicals if it better describes associations with effects or bioaccumulation. TOC-normalized values are calculated by dividing the dw concentrations by the fraction of TOC in the sample.

An *ex-situ* porewater passive sampler study was conducted on sediment samples collected from the 30 SQT stations to measure concentrations of a subset of chemicals. Measured concentrations within the sampler extracts were converted to equilibrium concentrations as described in the *Sediment Quality Triad and Porewater Data Report* (Tierra 2016d). These calculated porewater equilibrium concentrations will be used in the evaluations. Polychaete (*N. virens*) tissue chemistry data, reported on a wet weight (ww) basis, will be used for bioaccumulation modeling. Lipid normalization may be used if it better explains the relationship to sediment chemistry. Lipid-normalized values are calculated by dividing the ww concentrations by the fraction of lipids in the sample.

Sediment toxicity endpoints and BIC data will be used in the statistical evaluations to examine associations with sediment chemistry, and in the SQT scoring evaluation. Data handling and reduction methods for the BIC and sediment toxicity datasets are described below.

3.2 Benthic Invertebrate Community Data Reduction

Benthic invertebrate communities were characterized by identifying and enumerating the benthic invertebrates from sediment grab samples (three replicates per station) collected at the 30 SQT stations. The BIC samples were identified to the lowest practicable taxonomic level, in a manner consistent with other surveys performed in the New York/New Jersey (NY/NJ) Harbor Estuary (Weisberg et al. 1998). The following five standardized BIC metrics were calculated as described and presented in the *Sediment Quality Triad and Porewater Data Report* (Tierra 2016d):

1. Abundance per square meter (m^2)
2. Richness (i.e., number of species)
3. Shannon-Weiner diversity index (Shannon 1948)
4. Pielou's evenness index (Pielou 1966)
5. Swartz's dominance index (Swartz et al. 1985).

In addition, a multimetric benthic index of biotic integrity (B-IBI) was calculated using a method derived for the saline portions of the NY/NJ Harbor Estuary (Weisberg et al. 1998). Threshold values of the B-IBI for the classification of impacts are as follows: 1 to <2 = impacted, 2 to <3 = slightly impacted, and 3 to 5 = unimpacted (Weisberg et al. 1998). These six BIC metrics will be used in both the statistical evaluation and the SQT evaluation.

3.3 Sediment Toxicity Endpoints and Data Reduction

Ten-day acute and 28-day chronic toxicity tests were conducted on sediment from each of the 30 SQT stations using the amphipod *L. plumulosus* as described in the *Sediment Quality Triad and Porewater Data Report* (Tierra 2016d) and in accordance with appropriate USEPA guidance (USEPA 1994, 2001). The following endpoints were measured for each sample:

1. 10-day survival
2. 28-day survival
3. Growth rate (milligrams per organism per day) at Day 28
4. Reproduction (young per surviving adult) at Day 28.

Response data from each sediment sample were compared to laboratory control response data to determine if the sample response was statistically significantly different from control response (Tierra 2016d) using the program ToxCalc V 5.0.23[1]. For the SQT evaluation, response data will be converted to control-normalized response by dividing the observed response in the sample by the observed response in the control sample for each endpoint.

4. SEDIMENT QUALITY TRIAD EVALUATION

An SQT approach will be used to classify each of the 30 SQT stations with respect to degree of potential impact. Three categories (i.e., the triad) will be evaluated: BIC structure, sediment toxicity, and sediment chemistry. Each category will have equal weighting in the evaluation and will be assigned a score ranging from 0 to 1 (most impacted). The scores for each category will be summed (maximum possible score = 3) and the final score will be used to categorize the station with respect to degree of impact (Table 1). This methodology is consistent with the LPRSA BERA evaluation (Windward 2016) as directed by the USEPA (comment 71, USEPA 2015). The following subsections discuss how the scoring will be conducted within each of the three categories.

4.1 Benthic Invertebrate Community

The BIC will be evaluated using the reference envelope approach suggested by USEPA (2015). The five standardized benthic community metrics (abundance per m^2, richness [i.e., number of species], Shannon-Weiner diversity index, Pielou's evenness index, and Swartz's dominance index) are expected to decline in value as a result of physical and/or chemical degradation. Therefore, these values will be compared to the 5^{th} percentile of the reference distribution. BIC metrics that are equal to or greater than the 5^{th} percentile of the reference distribution will be considered comparable to reference and not impacted. BIC metrics that are less than the 5^{th} percentile of the reference distribution will be considered impacted. The reference site chosen for this evaluation is the Jamaica Bay Estuary, New York, from which SQT data

[1] The ToxCalc software tests the distribution of the response data (represented by response data from each of the five replicate beakers) and selects the appropriate parametric or non-parametric hypothesis test to compare the sample mean to the laboratory control mean. A resulting p-value less than 0.05 indicates that the sample mean is statistically significantly different than the laboratory control.

were collected under the REMAP program (USEPA 2003). Jamaica Bay is physically similar to the open water habitats of the NBSA and a sediment screening evaluation of this dataset was conducted for the LPRSA BERA (Windward 2016) as requested by the USEPA (2015). The Jamaica Bay sediment chemistry data were compared to effects range-low (ERL) and effects range-median (ERM) values (Long and Morgan 1990). Acceptable Jamaica Bay reference locations had three or fewer exceedances of ERLs and no exceedances of ERMs across all chemicals for which those sediment guidelines were available. In addition to meeting the chemical criteria, acceptable reference locations in Jamaica Bay were required to meet sediment toxicity criteria. Specifically, *A. abdita* survival results at Jamaica Bay reference locations were required to be ≥ 80 percent of the respective negative control response. The screening evaluation left a total of 25 samples from Jamaica Bay suitable for use as reference data. Table 2 presents summary statistics for the benthic metrics and the 5th percentile of the Jamaica Bay data to which the Newark Bay BIC metrics will be compared.

The value of the sixth metric, B-IBI, will be evaluated according to Weisberg et al. (1998) with respect to degree of impact. Each metric will be weighted equally (0.167) and the weights will be multiplied by the score, such that the maximum possible score for the BIC is 1, representing the greatest impact.

Each of the six BIC metrics will be evaluated as described above and scored from 0 (no impact) to 1 (impact). Each score will be equally weighted (0.167) so that the maximum possible score for this category is 1 (Table 1). For the five standardized BIC metrics, a score of 1 will be assigned if the value is less than the 5th percentile of reference data (Table 1); otherwise, a score of 0 will be assigned. The B-IBI will be scored based on the value of the index and according to the categories suggested by the authors (Weisberg et al. 1998). Unimpacted (B-IBI$\geq$3) stations will receive a score of 0. Slightly impacted (2$\leq$B-IBI<3) stations will receive a score of 0.5. Impacted (B-IBI<2) stations will receive a score of 1.

4.2 Sediment Toxicity

The sediment toxicity data comprise survival and sublethal (i.e., growth and reproduction) endpoints. Criteria for assigning toxicity are well established for the survival endpoints but not for the sublethal endpoints. Criteria used for the REMAP program (USEPA 2003) defined samples as toxic if the 10-day control-normalized percent survival was less than 80 percent and highly toxic if it was less than 60 percent. The 80 percent decision criterion is consistent with the criterion in the testing manual for discharge to U.S. waters (USEPA and U.S. Army Corps of Engineers [USACE] 1998) and was also suggested by Kennedy et al. (2009) for 28-day toxicity tests. Therefore, these criteria will be adopted to define toxicity of sediments from the NBSA based on the survival endpoints.

The sublethal endpoints are subject to considerable response variability compared to the survival endpoints. Eickhoff et al. (2014) subjected five control sediments to 28-day tests with *L. plumulosus* and found little variation in survival among the five control samples, which ranged from 96 to 99 percent. Within-batch replicate variance was also low for survival, with a mean coefficient of variation (CV) of 5.6 percent. However, growth and reproduction endpoints were more variable both among control samples and within batch replicates. Mean growth in control samples ranged from 0.7 to 1.8 milligrams per organism and the mean CV among replicates was 30 percent. Reproduction had a 10-fold range within control sediments (0.8 to 8.4 juveniles per surviving adult) and the mean CV among replicates was 79 percent. Kennedy et al. (2009) found significant variability in control sediment for growth (CV = 56 percent) and reproduction (CV = 63 percent), as well as significant variability in an interlaboratory comparison of

the growth and reproduction endpoints. *L. plumulosus* reproduction is known to be highly dependent on grain size (McDonald et al. 2010). This fact was illustrated in Kennedy et al. (2009) where it was found that a clean reference sample, which was relatively coarse-grained, had a very low response for the reproductive endpoint compared to the control (i.e., less than 10% of the control response). Based on the typical criteria used to define toxicity for survival endpoints (i.e., less than 80% of the control response) the clean reference sample would be classified as toxic based on the reproductive endpoint.

Given the large variability that can occur among control sediments for these sublethal endpoints and because there are no established control acceptability criteria for these sublethal endpoints, a simple statistical comparison to control sediment cannot be used to judge a sediment sample as toxic. Further, high variability makes default application of decision criteria (e.g., 20 percent reduction compared to control) inappropriate (McDonald et al. 2010). Therefore, for the SQT evaluation, it is proposed that toxicity based on sublethal endpoints will be determined if the sample response is statistically significantly different from control and less than 70 percent of control response. The sample will be presumed highly toxic if it is statistically significantly different from control and less than 50 percent of control response.

Scoring for the four sediment toxicity endpoints will be 0 for no impact, 0.5 for toxic samples, and 1 for highly toxic samples as described above and summarized in Table 1.

4.3 Sediment Chemistry

Sediment chemistry was measured in two ways: whole sediment concentrations and porewater concentrations. Each of these measures will be evaluated and given equal weighting (0.5) for this category (Table 1).

4.3.1 Whole Sediment Chemistry

Whole sediment concentrations will be compared to toxicity thresholds associated with predicted toxicity to amphipods (USEPA 2005). Chemical concentrations corresponding to a 20, 50, and 80 percent probability of observing sediment toxicity, defined as T20, T50, and T80, respectively, were developed for 37 chemicals using logistic regression modeling and a large database of sediment chemistry and toxicity data for marine amphipods from North American locations (USEPA 2005). The T20 and T50 values will be used to score the sediment concentration data (Table 1) from each station as described by the USEPA (2015) and as implemented in the LPRSA BERA (Windward 2016) as follows:

- If a sample does not exceed a T20 for any of the 37 chemicals, the score will be 0 (no impact).
- If a sample exceeds the T20 for any of the 37 chemicals but is less than or equal to the T50 for all chemicals, the sample will receive a score of 0.5.
- If a sample exceeds the T50 for any of the 37 chemicals, it will receive a score of 1

The T20 and T50 values are provided in Table 3. While not currently proposed for use in scoring the sediment chemistry concentrations, the T80 values are also provided in Table 3 for reference. The CPG conducted a reliability analysis of the T50 exceedance criterion for the LPRSA (Windward 2016) and found that the T50 had low specificity (i.e., high false positive rate) for predicting toxicity. The reliability of all three of these thresholds to correctly predict toxicity in the NBSA (i.e., T20, T50, and T80) will be

evaluated. If the T20 and T50 values are found to over- or under-predict the observed toxicity in the NBSA, an alternative scoring scheme for the NBSA sediment chemistry data may be proposed.

4.3.2 Porewater Chemistry

Porewater concentrations will be compared to the New Jersey Surface Water Quality Standards presented in the New Jersey Administrative Code, re-adopted October 17, 2016 (N.J.A.C. 7:9B), for the protection of aquatic life in saline waters and/or the USEPA aquatic ambient water quality criteria (AWQC) developed pursuant to Section 304(a) of the Clean Water Act. The New Jersey criteria will take precedence over the federal criteria. Federal criteria will be used where specific criteria have not been adopted by New Jersey. If the porewater concentration does not exceed any of the chronic criteria for the protection of aquatic life in saline waters, the sample will be given a score of 0. If the porewater concentration exceeds the acute criteria for the protection of aquatic life in saline waters, the sample will be given a score of 1. If chronic criteria are exceeded and acute criteria are not exceeded, the sample will be given a score of 0.5 (Table 1).

4.4 Sediment Quality Triad Categorization

The scores assigned to each metric will be multiplied by the assigned weight (Table 1), resulting in a maximum score of 1 for each category and a maximum score of 3 for each station. Each station will be classified as described by USEPA (2015) based on the total SQT score:

- No impact: $\leq$0.75
- Low impact: >0.75 and $\leq$1.5
- Medium impact >1.5 and $\leq$2.25
- High impact >2.25.

5. STATISTICAL EVALUATION OF SEDIMENT CHEMISTRY AND EFFECTS MEASURES

This section describes the statistical methods to be used to evaluate the association of sediment chemistry, both whole sediment concentrations and porewater concentrations, and measures of effects. The effects measures to be evaluated are the four *L. plumulosus* toxicity endpoints (10-day survival, 28-day survival, growth rate, and reproduction) and the six BIC metrics (abundance, richness, Shannon-Weiner diversity index, Pielou's evenness index, Swartz's dominance index, and B-IBI). In addition to chemical concentrations, physical parameters and characteristics of the sediment will also be evaluated (e.g., grain size, TOC, pH, acid volatile sulfide/simultaneously extracted metals, and habitat [intertidal vs. subtidal]). Community ordination techniques may be employed to better understand the effect of sediment chemistry concentrations on community structure. The goal of this evaluation will be to determine the role of chemical stressors on the benthic invertebrate community and to determine which chemicals are most associated with effects. If strong correlations are found with specific chemicals, site specific toxicity thresholds will be developed. The approach is summarized in Figure 1 and described in more detail in the following.

5.1 Univariate Analysis

A Spearman correlation matrix of effects measures vs. chemical (sediment and porewater) and physical parameters will be prepared and the correlation coefficients (cc) and associated p-values will be reported. All detected chemicals will be evaluated and the p-values will not be adjusted in any way (e.g., to reduce the Type I error rate). As described in the SQT QAPP (Tierra 2015c) the sample size for the SQT program was designed to detect a significant cc when the true cc is 0.5 or greater, with 95% confidence and 80% power. For a sample size of 30, any cc greater than 0.3 will be statistically significant.

Any chemical for which the cc exceeds 0.3 for *L. plumulosus* survival and/or at least two effects measures will be considered for additional consideration in multivariate analyses. It is possible that many chemicals will be correlated to at least two or more effects measures. Therefore, a screening process will be conducted in which chemical concentrations are compared to screening criteria. For sediment concentrations, if all samples are below the T20 threshold (USEPA 2005), the chemical will be excluded from further evaluation. For porewater concentrations, if all samples are below the chronic criteria for the protection of aquatic life in saline waters, the chemical will be excluded from further evaluation. The remaining chemicals will be carried forward in multivariate analyses.

All toxicity endpoints and BIC metrics that are correlated with at least one chemical will also move forward in multivariate analyses.

5.2 Multivariate Analysis

5.2.1 General Linear Models

Generalized linear models (Nelder and Wedderburn 1972) will be used to further evaluate relationships between chemical and physical parameters and a subset of effects measures to be determined based on the outcome of the univariate analysis. Generalized linear models have the form:

$$g(y_i) = x^j{}_i\beta + \varepsilon_i$$

where:

y_i is the response variable for the ith observation

x_i is a vector of j covariates or explanatory variables for observation i

β is a vector of unknown coefficients to be estimated by a least squares fit to the data y_i.

ε_i are assumed to be independent, normal random variables with zero mean and constant variance.

$g(y_i)$ is a link function of the response variable used to meet the assumptions of the error variance and/or to map the range of the response variable onto the range of the linear equation $(-\infty, \infty)$.

The expected value of y_i, denoted by u_i is:

$$g(u_i) = x^j{}_i\beta$$

The response variables, y, to be evaluated are the toxicity and benthic metrics. The explanatory variables, x, are the chemical and physical parameters that are to be carried forward following the chemical

screening process described above and may also be subject to transformation (e.g., logarithm transformation) to meet the assumptions of the error variance.

For continuous response variables, a traditional linear model will be evaluated where $g(u_i) = u_i$ (i.e., identity function). For binary response variables, such as survival, a logistic regression model will be evaluated where $g(u_i) = log(u_i/(1 - u_i))$.

5.2.2 Multicollinearity and Principal Component Analysis

If one or more explanatory variables are nearly linear combinations of others in the model (i.e., are highly correlated), the estimated coefficients (β) of the linear models are unstable and have high standard errors. This is quite likely for the set of physical and chemical parameters being evaluated in the NBSA dataset. If this is found to be the case, a principal component analysis (PCA) will be conducted with the set of explanatory variables (Morrison 1976).

The objective of PCA is to reduce the dimensionality of a dataset with a large number of correlated variables (e.g., chemical parameters). This reduction is achieved by transforming the data using matrix algebra to a new set of uncorrelated reference variables known as principal components (PCs), which are linear combinations of the original variables. PCs are sorted such that each, in turn, represents a smaller percentage of the variance within the dataset. The PCA will be conducted on a correlation matrix of the explanatory variables (i.e., normal standardization). PCA is subject to similar assumptions as linear modeling in that the residuals must be independent, normal random variables with zero mean and constant variance. Therefore, a log transformation of the variables may be conducted prior to data standardization.

Ideally, the multivariate structure will be largely and adequately explained by two or three PCs and each chemical will be highly correlated with a single PC, which aids in interpretation of the PCs. However, if this is not the case, an orthogonal rotation of the PCs (known as varimax rotation) will be used to create principal “factors” that are more easily interpreted. The goal of varimax rotation is to create factors to which variables are either highly correlated or highly uncorrelated. The resulting PCs or factors from this analysis can then be used in place of the original variables in the multivariate models described above. If the PCs or factors are found to be associated with effects, then the variables that are correlated with that PC or factor can be assumed to be associated with the effect; although it may not be possible to isolate a single chemical if it is more toxic than those with which it is correlated. Therefore, the outcome of the analysis may identify mixtures, not individual chemicals, that are associated with effects.

5.2.3 Community Ordination

Benthic metrics are useful indicators of BIC health, but some information is lost with the data reduction. Therefore, the BIC data may also be evaluated using community ordination. Ordination refers to a variety of techniques used to arrange benthic samples in relation to one or more coordinate axes and to each other to provide information about their ecological similarities. Community abundance data collected over environmental gradients are typically non-linear and are better suited to non-linear methods of analysis such as non-metric multidimensional scaling (NMDS) of a Bray-Curtis dissimilarity matrix (Ludwig and Reynolds 1988). Like PCA, NMDS is a method used to reduce the dimensionality of the community data, the result of which is that each benthic sample can be scored in the new smaller coordinate space. These scores can be plotted on graphs where samples near each other are judged more similar than those further away on the graph. Spatial patterns may be discernable if the BIC varies based on geographic or

geomorphic location. The resulting dimensions can be regressed against other variables such as individual chemicals, PCs, or factors to evaluate effects of stressors on changes in community structure.

5.2.4 Goodness of Fit Testing

Traditional multivariate linear models will be evaluated in a stepwise fashion and variables that are not significant (Wald test) or that do not add additional information or improve the adjusted coefficient of determination (R^2) value will be removed from the model.

Logistic regression models will be evaluated based on the adjusted R^2 proposed by Nagelkerke (1991).

5.3 Site-Specific Toxicity Threshold Modeling

If specific chemicals exhibit a strong correlation to toxicity endpoints, site-specific dose response modeling will be conducted. The model form will be selected as appropriate to the endpoint and could include: linear regression, logistic regression (survival), or a three-parameter non-linear log logistic model. The last is useful for control-adjusted endpoints that may not be constrained to a maximum of 1 as is the case with logistic regression. The three-parameter non-linear log logistic model has the following form:

$$y = Y_{max}/[1+(x/EC50)^{slope}]$$

where:

y	=	response value (percent of control response)
Y_{max}	=	regression-fitted maximum response (percent of control)
x	=	chemical concentration in sediment or porewater
EC50	=	50 percent effect concentration of the chemical being evaluated
slope	=	slope of the relationship.

Using the fitted models, the concentration at which the dose response curve is less than a prescribed percentage of the control response can be selected as a threshold effect level.

5.4 Bioaccumulation Modeling

A Pearson correlation matrix of log transformed sediment and polychaete tissue chemical concentrations will be prepared for all chemicals that were detected in the polychaete tissue. For any chemical in which the cc exceeds 0.7, a linear regression equation will be developed that describes the relationship and can be used to predict tissue concentrations based on other sediment data in the NBSA. TOC normalization and lipid normalization will be applied if found to reduce the error variance of the predictive models. The results of the polychaete tissue analysis will be used, in conjunction with the epibenthic clam and crab tissue data, to conduct the tissue-based exposure-risk evaluation under the BERA. This process will be described in the technical memorandum on the preliminary exposure factors and TRVs being developed for the NBSA by Arcadis.

6. SUMMARY

This memorandum summarizes the systematic and step-wise process that will be used to conduct the SQT assessment for the NBSA. This SQT represents a substantial portion of the risk characterization for benthic invertebrates for the NBSA BERA. The methods used are based on standard guidance and practice (e.g., USEPA 1994, 2001) and are consistent as possible with the SQT conducted for the LPRRP by the CPG (Windward 2016) in consultation with USEPA. The goal of the SQT assessment is to categorize levels of potential impact in various areas of the NBSA, and to quantify any exposure-response relationships between chemical contaminants in surface sediments and/or sediment porewater and toxicity endpoints and BIC metrics. A detailed SQT assessment will be prepared as an appendix to the NBSA BERA report, and the results will be incorporated into the risk characterization.

7. REFERENCES

Eickhoff, C., M. Grey, J. Pickard, S. Reimer, and D. MacDonald. 2014. The co-dependency of growth and reproduction: a closer look at the 28 day *Leptocheirus plumulosus* survival, growth and reproduction test. A presentation at the Society of Environmental Toxicology and Chemistry Annual Meeting, Vancouver BC. November 2014.

Kennedy, A.J., J.A. Steevens, G.R. Lotufo, and T.S. Bridges. 2009. A comparison of acute and chronic toxicity methods for marine sediments. *Marine Environmental Research* 68(3): 118-127.

Ludwig, J.A. and J.F. Reynolds. 1988. Statistical Ecology: A Primer on Methods and Computing. John Wiley and Sons, New York.

McDonald, B.G., C.A. McPherson, R. DeWynter, P. Kickham, and C. Brown. 2010. Incorporation of 28-d *Leptocheirus plumulosus* toxicity data in a sediment weight-of-evidence framework. *Ecotoxicology and Environmental Safety* 73:51-55.

Morrison, D.F. 1976. Multivariate Statistical Methods, Second Edition, New York: McGraw-Hill.

Nagelkerke, N.J.D. 1991. A Note on a General Definition of the Coefficient of Determination. *Biometrika* 78:691–692.

Nelder, J.A. and R.W.M. Wedderburn. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society*, Series A, 135:370–384.

Pielou, E.C. 1966. The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology* 13:131-144.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell Systems Technical Journal* 27(July and October):379-423 and 623-656.

Swartz, R.C., D.W. Schultz, G.R. Ditsworth, W.A. DeBen, and F.A. Cole. 1985. Sediment toxicity, contamination, and macrobenthic communities near a large sewage outfall. In: Boyle TT, ed, Validation and predictability of laboratory methods for assessing the fate and effects of contaminants in aquatic ecosystems. STP 865. American Society for Testing and Materials, Philadelphia, PA, pp 152-175.

Tierra. 2011. Technical Memorandum. Revised Meeting Minutes – Baseline Human Health and Ecological Risk Assessment Workshop June 28-29, 2011. December 1.

Tierra. 2013. Final Newark Bay Study Area Problem Formulation. Baseline Human Health and Ecological Risk Assessment. June.

Tierra. 2014a. Newark Bay Study Area Crab and Clam Sampling and Analysis Quality Assurance Project Plan. Revision 3a. August.

Tierra. 2014b. Newark Bay Study Area Fish Sampling and Analysis Quality Assurance Project Plan. Revision 2. October.

Tierra. 2015a. Reconnaissance Survey Report. Newark Bay Study Area. Baseline Human Health and Ecological Risk Assessment. April.

Tierra. 2015b. Technical Memorandum. Risk Assessment Field Sampling and Analysis Program – Newark Bay Study Area. November 30, 2015.

Tierra. 2015c. Newark Bay Study Area Sediment Quality Triad and Porewater Sampling and Analysis Quality Assurance Project Plan. Revision 2. August.

Tierra. 2016a. Draft Environmental Sample Collection. Newark Bay Study Area. June.

Tierra. 2016b. Draft Crab and Clam Data Report, Newark Bay Study Area, Revision 0, June.

Tierra. 2016c. Draft Sediment Quality Triad and Porewater Field Report, Newark Bay Study Area, Revision 0. July.

Tierra. 2016d. Draft Sediment Quality Triad and Porewater Data Report, Newark Bay Study Area, Revision 0, July.

Tierra. 2016e. Draft Fish Data Report, Newark Bay Study Area, Revision 0, September.

USEPA. 1994. Methods for Measuring the Toxicity of sediment-associated contaminants with estuarine and marine amphipods. EPA-600/R-94/025. Narragansett, RI.

USEPA. 2001. Methods for Assessing the Chronic Toxicity of Marine and Estuarine Sediment-associated Contaminants with the Amphipod *Leptocheirus plumulosus*. First Edition. EPA/600/R-01/020. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

USEPA. 2003. Final Report: Sediment Quality of the NY/NJ Harbor System: A 5-Year Revisit 1993/4 – 1998. An Investigation under the Regional Environmental Monitoring and Assessment Program (REMAP). EPA/902-R-03-002. USEPA-Region 2. Edison, NJ.

USEPA. 2005. Predicting toxicity to amphipods from sediment chemistry. United States Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC; EPA/600/R-04/030.

USEPA. 2015. EPA Response to CPG's September 2015 Response to EPA Comments Submitted May 1, 2015. Draft Lower Passaic River Study Area Baseline Ecological Risk Assessment Submitted June 13, 2014 by the Lower Passaic River Study Area Cooperating Parties Group. December.

USEPA and Battelle. 2015. Technical Memorandum, Lower Passaic River Restoration Project, Guidance for Revising the Sediment Quality Triad (SQT) Analysis. Norm Richardson (Battelle) and Chuck Nace (USEPA). December 11.

USEPA and USACE. 1998. Evaluation of Material Proposed for Discharge to Waters of the US – Testing Manual (Inland Testing Manual). EPA/823/B-98/004. United States Environmental Protection Agency, Washington, DC.

Weisberg, S.B., J.A. Ranasinghe, J.S. O'Connor, and D.A. Adams. 1998. A benthic index of biotic integrity (B-IBI) for the New York/New Jersey Harbor. Appendix C in Adams, D.A., O'Connor, J.S., Weisberg, S.B. 1998. Sediment quality of the NY/NJ Harbor System. EPA/902-R-98-001. United States Environmental Protection Agency, Region 2, Edison, NJ.

Windward. 2016. Lower Passaic River Study Area Baseline Ecological Risk Assessment: Revised Draft. October 7.

Table 1. SQT Scoring

Category	Metric	Weight	Scoring
Benthic Invertebrate Community	Abundance (per m^2)	0.167	0 = Within reference envelope ($\geq 5^{th}$ percentile of Jamaica Bay reference dataset) 1 = Outside reference envelope ($<5^{th}$ percentile of Jamaica Bay reference dataset)
	Richness	0.167	
	Shannon-Weiner diversity index	0.167	
	Pielou's evenness index	0.167	
	Swartz's dominance index	0.167	
	B-IBI score	0.167	0 = unimpacted (B-IBI≥3); 0.5 = slightly impacted (2≤B-IBI<3); 1 = impacted (B-IBI<2)
Sediment Toxicity *Leptocheirus plumulosus*	10-day survival	0.25	0 = control-normalized survival ≥80% and/or not significantly different from control 0.5 = control-normalized survival <80% and ≥60% and significantly different from control 1 = control-normalized survival <60% and significantly different from control
	28-day survival	0.25	
	Growth rate	0.25	0 = control-normalized response ≥70% or not significantly different from control 0.5 = control-normalized response <70% and ≥50% and significantly different from control 1 = control-normalized response <50% and significantly different from control
	Reproduction	0.25	
Sediment Chemistry	Sediment concentration[1]	0.5	0 = ≤T20; 0.5 = >T20 and ≤T50; 1 = >T50
	Porewater concentration[2]	0.5	0 = <chronic water quality criteria; 0.5 = >chronic water quality criteria; 1 = >acute water quality criteria

Notes:

[1]The notation Tp, such as T50, is used to denote the concentration at which "p" percent of samples are expected to be toxic, such as the probability that 50 percent of the samples would be toxic (USEPA 2005).

[2]Porewater concentrations will be compared to the New Jersey Surface Water Quality Standards presented in the New Jersey Administrative Code, re-adopted October 17, 2016 (N.J.A.C. 7:9B), for the protection of aquatic life in saline waters and/or the USEPA aquatic ambient water quality criteria developed pursuant to Section 304(a) of the Clean Water Act.

Table 2. Benthic Invertebrate Community Metrics in the Jamaica Bay Estuary Reference Dataset

Statistic	Abundance (per m^2)	Taxa Richness (No. of Taxa)	Shannon-Weiner Diversity Index	Pielou's Evenness Index	Swartz's Dominance Index
Sample size	25	25	25	25	25
Minimum	675	6.5	0.714	0.232	0.808
Maximum	146,750	47	2.82	0.855	7.76
Mean	25,683	26	1.78	0.575	3.31
Standard deviation	31,223	11	0.447	0.144	1.64
5^{th} percentile	**873**	**11**	**1.08**	**0.41**	**1.33**

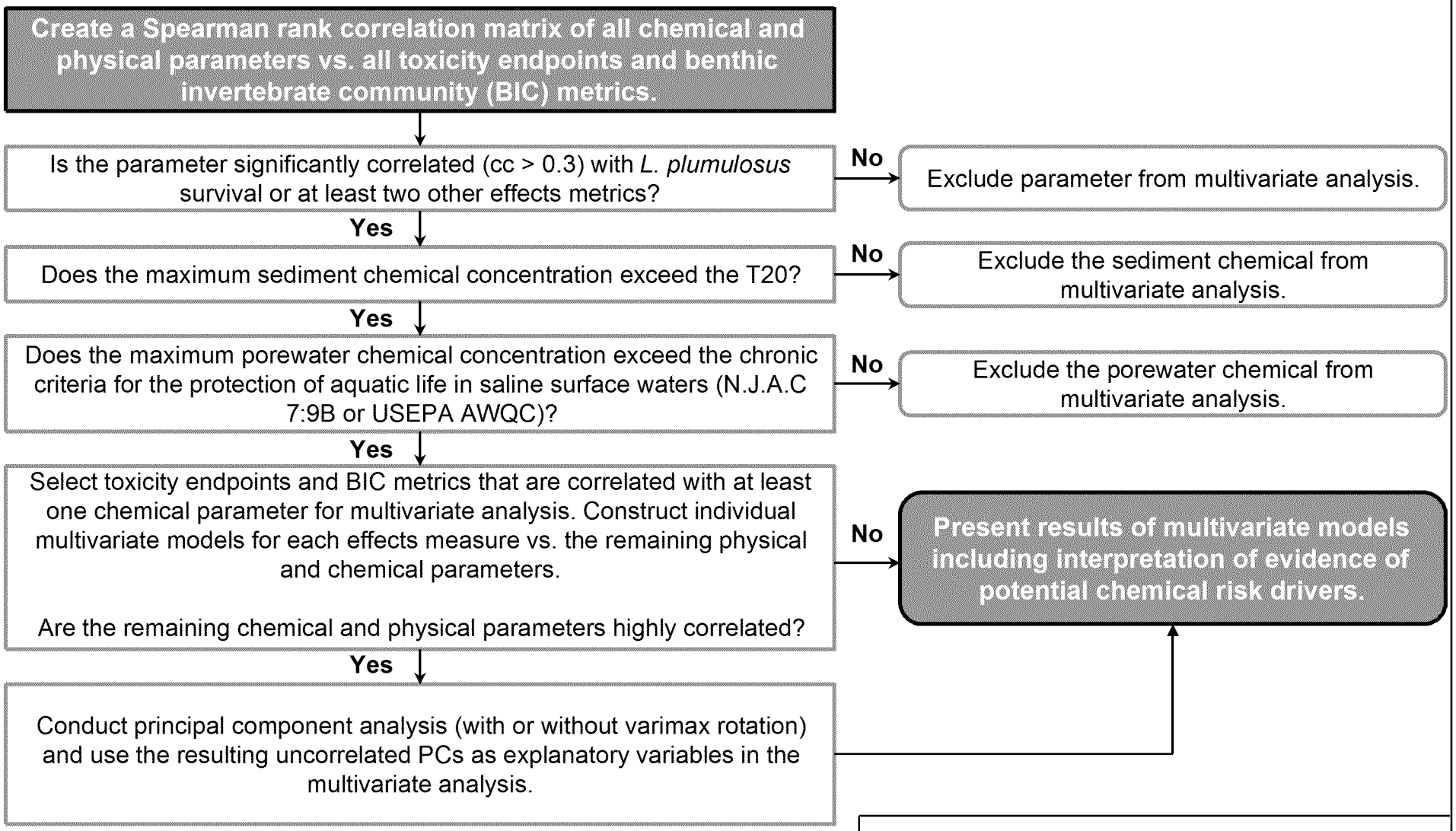
Source: From Table 2-4 in Appendix P of the LPRSA BERA (Windward 2016)

Table 3. Threshold Values to be used for the Scoring of Sediment Concentrations at SQT Stations in the NBSA

Chemical	T20	T50	T80
Metals (mg/kg dry wt.)			
Antimony	0.63	2.4	8.9
Arsenic	7.4	20	56
Cadmium	0.38	1.4	4.9
Chromium, total	49	140	410
Copper	32	94	280
Lead	30	94	300
Mercury	0.14	0.48	1.7
Nickel	15	47	150
Silver	0.23	1.1	5.8
Zinc	94	240	640
Polycyclic aromatic hydrocarbons (µg/kg dry wt.)			
1-Methylnaphthalene	21	94	430
1-Methylphenanthrene	18	110	700
2,6-Dimethylnaphthalene	25	130	710
2-Methylnaphthalene	21	130	770
Acenaphthene	19	120	710
Acenaphthylene	14	140	1420
Anthracene	34	290	2490
Benz(a)anthracene	61	470	3530
Benzo(a)pyrene	69	520	3910
Benzo(b)fluoranthene	130	1110	9410
Benzo(g,h,i)perylene	67	500	3710
Benzo(k)fluoranthene	70	540	4120
Biphenyl	17	73	310
Chrysene	82	650	5190
Dibenz(a,h)anthracene	19	110	690
Fluoranthene	120	1030	8950
Fluorene	19	110	660
Indeno(1,2,3-c,d)pyrene	68	490	3480
Naphthalene	30	220	1570
Perylene	74	450	2770

Chemical	T20	T50	T80
Phenanthrene	68	460	3060
Pyrene	120	930	6980
Polychlorinated biphenyls (µg/kg dry wt.)			
PCBs, total	35	370	3930
Organochlorine pesticides (µg/kg dry wt.)			
Dieldrin	0.83	2.9	10
p,p'-DDD	2.2	19	160
p,p'-DDE	3.1	100	3410
p,p'-DDT	1.7	11	76

Notes: Tp values reported in UEPA 2005 (Table 11). The notation Tp (e.g., T50) is used to denote the concentration at which "p" percent of samples are expected to be toxic (e.g., the probability that 50% of the samples would be toxic).



Notes:
cc = correlation coefficient
SQT = sediment quality triad
T20 = threshold for 20 percent probability of toxicity from: USEPA. 2005. Predicting toxicity to amphipods from sediment chemistry. United States Environmental Protection Agency, National Center for Environmental Assessment, Washington, DC; EPA/600/R-04/030.
N.J.A.C. 7:9B = New Jersey Administrative Code 7:9B, Surface Water Quality Standards.
USEPA AWQC = United States Environmental Protection Agency Ambient Water Quality Criteria, Section 304(a), Clean Water Act.

NEWARK BAY STUDY AREA
REMEDIAL INVESTIGATION
BASELINE ECOLOGICAL RISK ASSESSMENT

Approach for Statistical Evaluation of SQT Data

ARCADIS Design & Consultancy for natural and built assets

FIGURE
1